V0.6 CEOS GRID Interim Task Team Project Plan

Sept 5, 2002

Contributors : Yonsook Enloe, Dick desJardins, Jeff Smith, Allan Doyle, Glenn Rutledge, Stuart Doescher, Christophe Caspar, Ivan Petiteville, Pedro Gonzalves, Luigi Fusco, Debbie Ladwig, John Wilson

Table of Contents

1. Background

Many organizations and computer vendors have begun to research, implement and deploy Grid technologies and networks. Grid provides basic middleware services for seamless distributed computing and data management. Services such as resource request and allocation, security authorization and authentication, access control to high performance computing services, including specified qualities of service (QoS) from high performance networking services and storage area network (SAN) services, are all available through standard Grid APIs providing transparent access to distributed Grid infrastructure services. Networked Grid environments make it much easier for providers and users to implement and operate distributed enterprises, especially enterprises involving large amounts of distributed data and requiring user friendly but agency controlled support for access to multiple distributed computer models and computer processing power.

CEOS agencies have the mission to observe Earth from space and to provide data, information and models required for scientific investigations of Earth systems, particularly in support of the science priorities of the Integrated Global Observing Strategy (IGOS) partners. CEOS WGISS Subgroups determined at their Joint Meeting in Darmstadt in September 2001 that Grid technology was potentially important enough to the future of CEOS information systems and services that the Subgroups should investigate Grid technologies in a special workshop in 2002.

The special workshop was held in conjunction with the May 2002 Subgroups meeting in Frascati. As an outcome of the special workshop, the Subgroups decided to establish an Interim CEOS Grid Task Team to prepare a plan for CEOS Grid Prototyping Activities in 2003. The plan would be presented to WGISS-14 in Honolulu in September 2002 for review and approval.

The Interim CEOS Grid Task Team held workshops on June 26-28, 2002, in College Park, Maryland, and on September 9, 2002, in Alexandria, Virginia. This draft Plan is the output document from these workshops. Upon approval by WGISS, this document will become the CEOS Grid Prototyping Project Plan (version 1.0).

Current Roles and responsibilities (July 2002)

Interim Task Team Leader: Dick desJardins

Interim Task Team Deputy Leader/Project Plan Editor: Yonsook Enloe

Applications Lead: Luigi Fusco, ESA.

Engineering and Technology Lead: Allan Doyle

Network Lead: Jeff Smith

2. Introduction to the CEOS GRID Prototyping Project

Although the Subgroups want to investigate the application of GRID technologies to CEOS agencies information systems and applications by forming a task team, at the June Interim GRID task team workshop team members recognized that the CEOS agencies need to build up expertise in GRID technologies before tackling agency driven application projects. Each participating CEOS agency has at least one application project with multiple partners (internal and external to CEOS) which GRID technologies could potentially benefit. Therefore, team members agreed to a two phase approach during the first year of the task team activation.

The first phase of the first year is a GRID core technology exploration period where team members learn about the GRID technology and learn how to apply the technology to their applications. GRID experts from existing GRIDS will be assisting in this technology learning phase.

In the second phase of the first year, the team members will lead the execution of their specific application projects with their potential application partners. The agency specific application projects are described in the appendices. Other CEOS agencies are encouraged to participate in any applications projects of interest.

During the second year, additional CEOS partners who wish to join the CEOS GRID will be encouraged to do so. They may be inspired to join the CEOS GRID after viewing demonstrations of the application project prototypes. This will encourage additional members as well additional CEOS collaboration. Also during the second year, discussions on potential uses of GRID technologies in the different WGISS Test Facilities are expected.

This project plan only covers the first year. The GRID Task Team is expected to revise the project plan as plans for the second year develop.

2.1 Objectives of CEOS Grid Prototyping Project

To investigate the applicability of Grid technologies to CEOS needs, WGISS should establish a CEOS Grid Prototyping Project to pursue two objectives in 2003, with the aim of preparing to pursue a follow-on third objective in 2004, as follows:

Objective 1 (First Half-year, October 2002 to March 2003): Establish a CEOS Grid Technology Core Testbed with at least three major participating agency nodes. The purpose of the Grid Technology Core is to establish an immediate Grid capability base (including technologies, pilot applications, and knowledgeable people) within the participating CEOS agencies.

A Mid-year Review and Replan would be carried out in March or April 2003 to certify progress along the "learning curve" for Grid technologies and services, and to assure realism of the second half-year plan and success criteria for demonstrating Grid-enabled applications.

Objective 2 (Second Half-year, April 2003 to September 2003): Demonstrate at least three CEOS Grid-enabled Applications, each involving at least one CEOS agency and partner site which could include other CEOS agencies. The purpose of the Grid-enabled Application Demonstrations is to show proof of concept, evaluate benefits, and obtain lessons learned, from infusion of Grid technologies from the Technology Core into real CEOS agency information systems and applications.

A report on Project first year results, including demonstrations of the CEOS Grid Applications, would be presented to WGISS in Fall 2003. If the first year is deemed successful and WGISS gives the go-ahead for the follow-on year, the following objective would be pursued in 2004.

Objective 3 (Second Year, October 2003 to September 2004): Infuse applicable Grid technologies into member agency information systems and into at least one WGISS Test Facility (WTF). The outcome of this Grid technology infusion would be to create a persistent CEOS Grid that would be available to support future CEOS agency initiatives.

2.2 Schedule for CEOS Grid Prototyping Project

The CEOS Grid Prototyping Plan is based on a view of activities and phases needed to achieve the above objectives, as depicted in Figure 1.
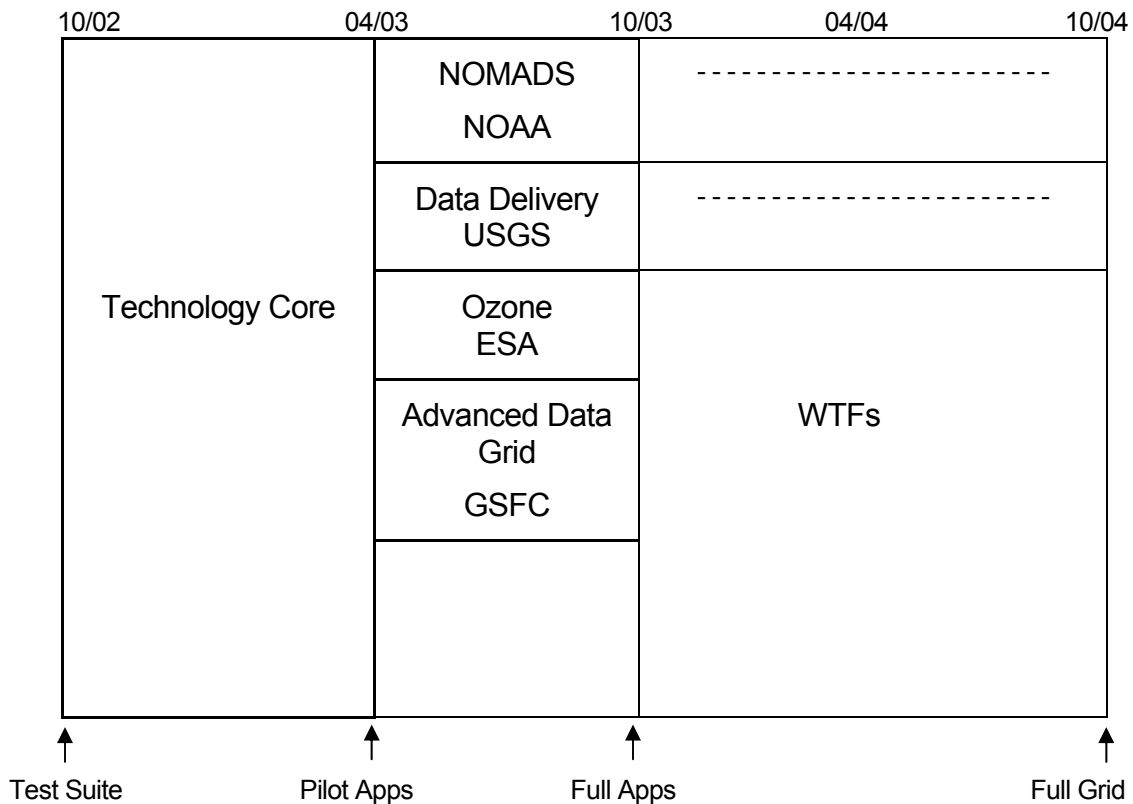
	10/02		04/03		10/03		04/04		10/04

Technology Core	NOMADS NOAA	- -
	Data Delivery USGS	- -
	Ozone ESA	WTFs
	Advanced Data Grid GSFC	

↑ Test Suite   ↑ Pilot Apps   ↑ Full Apps   ↑ Full Grid

Figure 1 : CEOS Grid Prototyping Project Schedule

The Prototyping Plan is divided into three main phases, corresponding to the three Project Objectives. In Phase 1 (First Half-year, October 2002 to March 2003), the participating agencies will establish an Initial CEOS Grid Technology Core Testbed, as depicted in Figure 2. Establishment of the testbed would include creation of a Grid software and services test suite and implementation of pilot versions of all Phase 2 applications. This would achieve Objective 1.
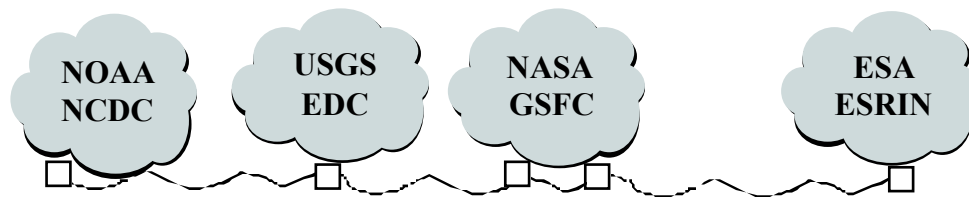
Figure 2: CEOS Grid Technology Core Testbed

In Figure 2, note that the testbed nodes are shown outside the agency campus firewalls. This approach would enable quick implementation of the testbed. As depicted in the figure, some agencies might want to implement more than one node on the testbed. The network "wiring diagram" shown in Figure 2 is only figurative; the network architecture and topology needed to support the Grid Technology Core Testbed will be worked out by the Network Task Team during the first three months of Phase 1.

The Phase 1 (First Half-year, October 2002 to March 2003) schedule allows the first three months (Phase 1a, October to December 2002) for setting up and interconnecting the machines, downloading Globus and other software, handing out CEOS Testbed certificates (good within CEOS Grid Prototyping sites only), and helping CEOS technical people at each site become knowledgeable about Grid software, services and features.

During the second three months of Phase 1 (Phase 1b, January to March 2003), each Application to be demonstrated in Phase 2 will be implemented first in a sample or pilot application subset installed in the Technology Core Testbed. The purpose of these pilot application subset installations is to shake down the Grid software and services for each application. By focusing on pilot subsets, the applications can be demonstrated in limited formats within a fully supportive Grid environment, in which all external variables are under Testbed control.

Phase 1 will be concluded with a Mid-year Review and Replan activity in March or April 2003. This will allow the participating agencies and outside Grid experts to certify adequate progress along the "learning curve" of Grid technologies and services, and to assure realism of the plan and success criteria for demonstrating Phase 2 Grid-enabled applications by the end of the Project First Year. This review would be essential because climbing the Grid learning curve is an essential progress metric for Phase 1, yet the schedule of activities leading to achievement of the initial Technology Core interoperability test suite can only be "guesstimated" for this Project Plan. Indeed, this is the very reason for carrying out this prototyping in the three well-defined phases defined in this Plan. The replan would be essential because only after building the Technology Core testbed and implementing the pilot applications will the nature of, and criteria for, Grid-enabled application "success" by the end of the Project First Year be well understood.

In Phase 2 (Second Half-year, April to September 2003), the Grid-enabled pilot Applications will be extended outside of the Technology Core Testbed and installed into real agency and supporting Grid information systems. Three Grid-enabled applications – e.g., NOMADS (NOAA), Data Delivery (EDC), and Ozone (ESA) – will be demonstrated and evaluated by the CEOS Grid Task Team, participating agencies, and supporting external Grids. This will achieve Objective 2, showing feasibility, evaluating

benefits, and obtaining lessons learned, from infusion of Grid technologies from the Technology Core into real CEOS agency information systems and applications.

Towards the end of Phase 2, the Task Team will prepare a detailed Project Plan for a follow-on Phase 3 (Second year, October 2003 to September 2004), including objectives, approach and expected outcomes. If WGISS gives the go-ahead for Phase 3, then the Task Team would provide technical support and expertise to enable WGISS to establish a worldwide CEOS Grid including all major agencies, data providers and affiliated information systems, by the end of 2004. This would achieve Objective 3, creating a persistent CEOS Grid Environment that will be available to support future CEOS agency initiatives.

3. Engineering and Technical Plan

The Technical Team will focus on establishing (or extending/documenting) a base level of functionality at each participating organization which can be expanded as organizations gain experience with the Grid technologies.

Emphasis will be on coordination and communication so that lessons learned by team members can be recorded and made available to others. Central to this coordination role will be regular communications via teleconference and email as well as web resources such as lists of resources, short "how to" or "lessons learned" documents, status information, etc.

During the first phase the emphasis will be on getting at least one basic machine configuration running at each members' site, connecting these sites together in a CEOS Grid Virtual Organization, and initiating discussions with one or more existing Grids/Virtual Organizations about joining their Virtual Organizations.

Since different participants will be starting with different levels of existing expertise, will be able to progress at different rates, and will be able to muster different levels of resources, we will try to accommodate different levels of participation.

The CEOS Grid activity is expected to be a multi-year activity. During this time, participating organizations will be engaged in acquiring expertise at different rates and in settings that are not always directly connected to the CEOS WGISS activities. During these times, participating organizations will nonetheless share their experiences with the WGISS Grid activity participants and will be marshalling Grid-enabled resources that will eventually be brought to bear on a CEOS WGISS project, most likely a future WTF. A key measure of success during the initial learning phases will be the degree to which participating organizations will have advanced their own understanding of Grid technologies as a result of having participated in this activity.

3.1 Technical Coordination Activity

The Technical Team will carry out a coordination activity that will consist of at least the following items:

a.   Establish a technical working group.

The TWG will consist of at least one person from each organization who can participate actively in the teleconferences, email discussions, etc. Membership in the technical group is open but it is expected that each participating organization will dedicate one or two individuals who will maintain continuity of participation.

b.   Hold weekly teleconferences

The purpose is to provide status updates, allow for questions, discussion, etc. as well as brainstorming.

c.   Establish an information exchange web site

The web site will allow members to contribute any materials they feel is relevant to the work of the TWG. Currently, the web site is available at http://grid-tech.ceos.org/gridwiki and the web site is designed to be collaboratively expanded by participants as they develop materials for the group. Additional collaborative software can be installed as needed.

d.   Establish technical mailing list

This list will initially be the ceos-grid@ceos.org mailing list but we anticipate splitting off a new list as technical email volume rises.

e.   Establish list of resource commitments & status

The web site will contain sections where participating organizations can document their available resources and the status of those resources. Owners of the resources will be encouraged to keep their portions of this list up to date.

   i.   Machines & configurations

f.   Storage

g.   Data availability

h.   …

3.2 Site Configuration Activity

Participating members will each establish at least one machine, accessible to the other members with the following suggested configuration:

- High end Linux or other capable machine, minimum 500 GB.
- Globus 2.0 distribution, including MDS 2.1 and GridFTP.

Additionally, the following are desirable:

- Externally accessible site on high performance research and education network (HPREN).
- Ability to connect to agency or other community Grids.
- Agency applications of interest, either as champion or user.
- Some agency data sets offered to applications.

Initial configuration activities (at a minimum):

1. Install basic Globus toolkit

2. Announce status to technical working group

3. When additional sites are ready, begin connectivity testing

4. Work with the Network team to set up automated connectivity and performance testing

3.3 Certificate Management

Key to Grid activities is membership in a Virtual Organization (VO). This means a community of resource providers who all agree to accept each others credentials. These credentials take the form of X.509 certificates. Established Grids also have established rules about what kinds of credentials they will accept.

Part of our activity will involve selecting one or more specific, existing Grid Virtual Organizations to join, approaching those VOs, and negotiating access agreements with them. This activity is likely to take some time. As an interim measure, we will be either getting test certificates from an external certificate authority (CA) or we will establish a ceos-grid CA to generate these test certificates. Using the test certificates, we can perform initial experiments among ourselves.

Grid resources (i.e. machines running grid software) can participate in any number of VOs at any time. Thus, as we develop relationships with other VOs, we can tie into those without having to stop accepting the test certificates.

Initial certificate management activities will include:

1. Generate our own ceos-grid test certificates.

2. Approach an existing community (or more than one) with a request to join their virtual organization.

An important point is that any machine/resource/group can participate in more than one Grid or Virtual Organization. This means that participants can progress at their own rates, can join whatever VO they feel is advantageous, and can still help the rest of the CEOS Grid participants advance their capabilities.

4. Network Support Plan

Below is an outline showing how the Network Team will support the Interim Grid Task Team. The Network team will assist each site's Points of Contact (Overall POC, Network POC, Host POC) to fill in the site, host, network, and traffic sub-sections of section 1. Once section 1 is filled out, the Network team will start performing section 2 activities in support of each Grid application.

1. Information and support provided by Overall Point of Contact (OPOC) to the Network Team
    - Site (provided by OPOC)

- o Location: facility, org
- o Network Point of Contact (NPOC)
- o Host computer Point of Contact (HPOC)
- Host (provided by HPOC)
  - o Location: B??/Rm??
  - o Platform
  - o Operating system
  - o Memory
  - o Network Interface Card
  - o Account support
- Network (provided by NPOC)
  - o LAN
  - o WAN
- Traffic characteristics (provided by OPOC)
  - o Data rates
  - o Pattern (burst, period, dates)

2. Information and support provided by Network Team to the GRID Interim Task Team
   - o Network Performance Measurement
     - o Route stability
     - o Packet loss
     - o Bandwidth throughput
   - o Routing optimization
     - o Path verification
     - o Transit negotiation
   - o Results Reports
     - o Web
     - o Presentations
     - o Network maps
   - o Connectivity and performance status monitoring
     - o Network problem solving
     - o Routing optimization
     - o Host stack tuning
     - o Security Denial of Service
   - o New technology infusion (where applicable)
     - o QoS
     - o Etc.

Appendix A – USGS Application Plan

Purpose

Preliminary investigation is needed to explore feasibility of utilizing GRID technologies for the delivery and reception of data. Primarily this would employ the services of the GridFTP and certificate authority.

Title: Data delivery utilizing GridFTP.

Point(s) of contact:

Stuart Doescher, USGS/EDC, doescher@usgs.gov, (605)-594-6013

???

Application area and specific scenario.

Exploration of the GridFTP and certificate authority process will lead to a detailed understanding of their applicability to implement this technology. The focus would be on the aspects of data delivery to the scientific user community and with receiving data into the archive from producer/reception sites.

Historically, users have tended toward traditional data deliver on media. Network data delivery has been relegated to the scientific users that have a critical time need and have the appropriate network capabilities. However there continues to be an increase in the customers that desire ftp delivery of data. Currently, the means of enabling ftp data delivery are primitive with implementation of such capabilities as semi-anonymous ftp and very limited ftp push. Security is one of the major concerns associated with current ftp service mechanisms. GridFTP services may be able to improve the current ftp data delivery environment.

Likewise there is a potential opportunity to improve the delivery of the high volume data flows from the reception sites and/or production sites to the archives.

Technology exploration focused scenario.
Phase 1. (0 to 6 months)
1. Install GLOBUS toolkit.
2. Develop understanding of GridFTP
3. Identify Technology Core testbed site. (maybe local)
4. Tie into existing GRID certifying authority.
5. Explore usage.
6. Determine costs/benefits.

Phase 2. (6 to 12 months)
1. Install certification authority.
2. Reestablish Phase 1 test within the new certificate environment.
3. Determine costs/benefits.

Phase 3. (13 to 24 months)
1. Investigate CEOS Test Sites WGISS Test Facility requirements for applicability to Data GRID services
2. Determine cost/benefits
3. Pursue Working Group Cal/Val and WGISS concurrence for a pilot.

Updates will be forthcoming as the Cal/Val plan is refined.

Grid locations.

- LAI, VALERI, GOFC and EOS Core Sites initially

- CEOS Core test sites

- CEOP test sites in a later phase

- LTER test sites in a later phse

- GLCTS test sites in a later phase

- Joint Agency Coordination Commercial Imagery Evaluation (JACIE) test sites in a later phase?

Confirmed Datasets.

Potential Datasets

- MODIS Subsets

- ETM+

- SPOT-HRV

- SPOT-Vegetation

- AATSR

- POLDER

- MERIS

Grid services.

It is anticipated that the project will be utilizing GridFTP for data deliverables, Globus Replica Catalog for keeping track of replicated files and the Globus Replica Management for managing files in a high performance, wide area data storage environment.

Applications and data.

Application and data focus is on acquisition and evaluation of data from multiple sources, making Landsat data available for CalVal. Subsetting, reprojection, data fusion. CalVal potentially can use lower performance networks at some sites.

Appendix B – NOMADS Application Plan

Title : "The NOAA Operational Model Archive and Distribution System (NOMADS)". NOMADS is a distributed grid framework promoting standards across multiple institutions for access to climate and NWP models for analysis and intercomparison.

Point(s) of contact. Glenn K. Rutledge (Project PI), NOAA, National Climatic Data Center, Glenn.Rutledge@noaa.gov (828) 271-4097. Danny Brinegar Danny.Brinegar@noaa.gov.

Application area and specific scenario.

To address a growing need for remote access to high volume numerical weather prediction and global climate models and data, the National Climatic Data Center (NCDC), along with the National Centers for Environmental Prediction (NCEP) and the Geophysical Fluid Dynamics Laboratory (GFDL), initiated the NOAA Operational Model Archive and Distribution System (NOMADS) project. NOMADS addresses model data access needs as outlined in the U.S. Weather Research Program (USWRP) Implementation Plan for Research in Quantitative Precipitation Forecasting and Data Assimilation to "redeem practical value of research findings and facilitate their transfer into operations." The NOMADS framework was also developed to facilitate climate model and observational data inter-comparison issues as discussed in documents such as the Intergovernmental Panel on Climate Change (IPCC 1990, 1995, 2001) and the U.S. National Assessment (2000). NOMADS is being developed as "*A Unified Climate and Weather Archive*" so that users can make decisions about their specific needs on time scales from days (weather), to months (El Nino), to decades (global warming).

The Project

1. NOMADS is a network of data servers using established and emerging technologies to access and integrate model and other data stored in geographically distributed repositories in heterogeneous formats. NOMADS enables the sharing and inter-comparing of model results and is a major collaborative effort, spanning multiple Government agencies and academic institutions. The data available under the NOMADS framework include model input and Numerical Weather Prediction (NWP) gridded output from NCEP, and General Circulation Models (GCM) and simulations from GFDL and other leading institutions from around the world. The goals of NOMADS are to:

2. provide access to NWP and GCM's model output and provide the observational and model data assimilation products for Regional model initialization and forecast verification,

3. improve operational weather forecasts,

4. develop linkages between the research and operational modeling communities and fosters collaborations between the climate and weather modeling communities,

5. promote product development and collaborations within the geo-science communities (ocean, weather, and climate) and fosters inter-disciplinary research to study multiple earth systems using collections of distributed data under a sustainable system architecture.

The Partners

The NOMADS framework is actively partnering with existing US and development activities including the Comprehensive Large Array Stewardship System (CLASS); the National Oceanographic Partnership Program's (NOPP) National Virtual Ocean Data System (NVODS); the Department of Energy's Earth System Grid (ESG); and the Thematic Real-time Environmental Data Distributed Services (THREDDS) project being developed through the National Science Foundation and Unidata. To ensure that Agency and Institutional requirements are being met, the NOMADS collaborator's have established Science and Technical Advisory Teams. These newly established panels would ensure

the NOMADS system and metadata architecture could provide necessary inter-operability; and develop data archive requirement recommendations to NOAA.

NOMADS will make available to the CEOS-Grid, the core applications and subsets of available NOMADS data for access across the grid. This will initially include the Center for Ocean-Land-Atmosphere (COLA) developed GrADS-DODS Server (GDS) allowing access to hyper-slices of data and client-side aggregation of data across NOMADS participating host servers (see confirmed CEOS-Grid confirmed Datasets below). Follow-on applications include Lawrence Livermore National Laboratory (LLNL) Climate Data Analysis Tools (CDAT) that will allow for climate model intercomparison and the generation of quality control statistics on various model parameters; and the NOAA Pacific Marine Environmental Laboratory (PMEL) developed Live Access Server (LAS) and "Ferret" client side analysis tool. Both these follow-on applications would be accessible under the NOMADS collaboration with the DOE's Earth System Grid (ESG) at LLNL.

Grid Partners. NOMADS will also function under the ESG as determined by requirements developed between NOMADS and ESG, and between ESG and the newly established Natural Environment Research Council (NERC) DataGrid (United Kingdom).

Grid Experts. NOMADS will rely on Dean Williams (LLNL) the ESG – PI as its Grid Expert; and Danny Brinegar (NCDC) as it's local NCDC Grid administrator.

Grid locations. Under the NOMADS framework, selected NOMADS team members would contribute their holdings and expertise on the "grid". NOMADS participating sites will be grouped into primary and secondary tiers, and each collaborating NOMADS partner collaborating under CEOS-grid, will determine their tier level and thereby their own level of involvement.

Primary sites will install CEIS-Grid available gird toolkits and make available data and applications currently being developed under the NOMADS. Applications and associated data being made by each institution will be further subdivided into 1) confirmed and 2) potential. "Confirmed" data and applications will include those already being made available under NOAMDS. "Potential" data sets could be those in which participating institutions looking to resolve future expected high volume data transfer and processing requirements using these grid technologies requiring high bandwidth distributed technologies. Primary sites will initially include NCDC (Asheville), LLNL (Livermore), GFDL (Princeton), Lawrence Livermore National Laboratory (LLNL) (Livermore), and NCEP (Camp Springs). Additional NOMADS sites may include PMEL (Seattle), and CDC (Boulder). Institution specific tasks would have to be further coordinated based upon site mission and available resources.

Confirmed Datasets. Confirmed data sets available through NCDC for the primary tier will include a subset of the National Weather Service NOAAPort model output dataset being ingested at NCDC. These data include the mesoEta, RUC, and the Global Forecast System (GFS) in GRIB formats. Confirmed applications include the COLA developed GrADS-DODS Server.

NCDC: NCEP RUC Analysis and Fcst Fields;

NCEP: Real-time and retrospective GFS Data Assimilation Model Initialization and Restart Files;

NCAR: Subset of the Community Climate System Model (CCSM);

GFDL: R-30 DecCen Coupled Climate Experiments, Ocean Assimilation Experiments, and Input and Output Files for the Flexible Modeling System.

LLNL: AMIP Climate Model Projects (non-restricted); and AMIP / Probabilistic information

Potential Datasets. PMEL: tbd; CDC: tbd

Grid services. Basic Globus 2.0 services including GridFTP, Globus information services (MDS 2.1), need certificates.

Appendix C – ESA Ozone Application Plan

Title: Interaction with GoD-WM: GRID on Demand for Web Mapping

Point(s) of contact

Pedro Goncalves, ESA/ESRIN (pedro@inovagis.org) (+39) 06 941 80538

Christophe Caspar, ESA/ESRIN (Christophe.Caspar@esa.int)

Luigi Fusco, ESA/ESRIN, luigi.fusco@esa.int, (+39) 0694180539

Ivan Petiteville, ESA/ESRIN, Ivan.Petiteville@esa.int

Purpose

Develop a CEOS testbed for interfacing CEOS interoperability technologies with GRID environments. Demonstrate the integration of technologies for selected EO applications: the first selected case is for Ozone data handling and exploitation.

Application area and specific scenario

Atmospheric science projects generally require integration and validation of core EO and ground-based measurements, their assimilation in scientific models for an effective utilization of the derived geophysical parameters.

GoD-WM proposes the development of a Portal for Earth Science Applications Browser, compliant with in OpenGIS specifications, and integrating available Grid infrastructures (e.g. DataGrid, Earth System Grid), with CEOS Data Warehouse systems and with GIServer[1] Web Map Server (WMS). Starting with data available in EO data warehouses, the system will allow the execution of remote services within the GRID environment as a new layer of information that can be afterward integrated with other servers following the Web Map Specification (OpenGIS, 2001). More specifically, the GoD-WM portal will give the user the possibility of performing the following tasks within an Internet Browser:

6. Temporal and spatial selection of data to be processed

7. Data transfer from data warehouses in user space

8. Data transfer to and from the GRID storage elements

9. Job selection, launching and status information queries.

10. Result retrieval and visualization with other WMS layers

All these functionalities are to be presented in HTML user interface implemented using a client application with some generic functions developed in JavaScript.

The following technologies are considered in the project:
- Catalogue Interoperability (e.g. MUIS at ESA, IMS at NASA)
- Archive Data Management (e.g. AMS at ESA, …)
- Web Mapping tools (e.g. GIServer at inovaGIS, WEB MAPPING at ESA, …)
- On demand generation of level 2 GOME products based on GRID developed algorithms at ESA and at KNMI (NL)
- Validation of such products against ground Lidar measurements available at IPSL (F). Visualization of results.

---

[1] More information about the GIServer Web Map Server can be obtained in its on-line tutorial available at:
http://www.inovagis.org/giserver/tutorial/tutorial.htm

- Possible assimilation of final validated products in modeling tools running on GRID environment (to be confirmed)
- Visualization of model results (to be confirmed)

The main area of application is the Web Map data integration with GRID applications for ESA Ozone data and the NOMADS Application Plan. Nevertheless the CEOS-GRID partners might also consider the following parallel tasks accordingly to the demonstrated interest:

1. Visualization of GOME Level 2 data using the NASA processing tools for data visualization (TBD)
2. NOAA TOMS and ESA GOME Total Ozone comparison and integration (TBD)

Technology exploration focused scenario

Phase 1. (0 to 6 months)

1. Design and implementation of Client/Server SOAP interface wrapper for EDG GRID client for job construction and submission
2. Implementation of graphical SOAP clients in Linux (using JAVA and Kylix) and in Windows (using Delphi)
3. Test job submission for the Ozone application with pre-defined jobs.

Phase 2. (6 to 12 months)

1. Analysis of OpenGIS Web Map/Coverage/Feature and Styles Definition specifications for detailing interface input/output parameters required by the GRID jobs
2. Design of a Service Management Layer component for stateless requests (e.g. launching GRID jobs from OGC Web Services)
3. Design and implementation of Client/Server SOAP interface wrapper for MUIS client access (access to ESA Catalogue) – interface with CIP to be investigate as future development
4. Design and implementation of Client/Server SOAP interface wrapper for AMS client (access to ESA Archive Data Management) – interface with GRIDftp to be implemented as future development
5. Test data finding, retrieval and job submission for the Ozone Application

Phase 3. (12 to 24 months)

1. Design of the integration schema for the NOMADS application plan
2. Investigate NOMADS application requirements.
3. Installation of the necessary EDG middleware in the NOMADS application
4. Demonstration of integrated interface between NOMADS/EDG via web mapping. Application case test definition and implementation

Confirmed Datasets.

Potential Datasets.
The following EO mission data are considered of interest for the project:

- ESA GOME instrument data on ERS-2
- NOAA TOMS instrument data on various missions
- ESA SCIAMACHY, GOMOS, MIPAS instruments (tbd) on Envisat
- Lidar ground measurements acquired at different facilities

Grid locations.

At least the following GRID locations are supposed to be involved:

- Nodes in EDG CEOS Virtual Organization (all EDG EO Virtual Organisation sites until EGD is capable to handle different cluster of nodes for different VO's). In particular the reference node will be the ESA-ESRIN node (UI, CE, SE)

- ESG – Earth System Grid, … TBD

- Others (To be identified, e.g. UK e-science) …

Grid services.

All basic GRID services available within EU DataGrid (and within tbd GRID environments) will be used (certification/authentication, GRID FTP, GRAM, GSI, …).

In particular it will be interesting to access and validate the utilization of the UI (User Interface), CE (Computing Element) and SE (Storage Element), the RB (Resource Broker) and the RC (Replica Catalogue) for the specific identified application.

At present the EDC reference points (for the EO Virtual Organisation) are:

- CA: nationally provided. ESA uses the CNRS (F) Certificate Authority. The EO Virtual Organisation is managed by KNMI (NL)

- RB: at present we use CNAF (Bologna – I) and CERN (CH)

- RC: at present the EO VO replica catalogue is located in NIKHEF (NL)

Appendix E – NASA GSFC Advanced Data Grid Application Plan

Purpose

The objective of the GSFC Advanced Data Grid Prototype project is to demonstrate the execution of scientifically 'meaningful' applications software, e.g. a climate model, in a data grid environment that includes two 'collaborating' data systems, each of which employs the Storage Resource Broker (SRB)/Metadata Catalog (MCAT) and DataCutter to manage massive data sets: one, current and on-line or near-line; the other, aged and off-line.

Title: NASA GSFC Advanced Data Grid Prototype

Point(s) of Contact:

Debra Ladwig, NASA,/GSFC, debbie.ladwig@gsfc.nasa.gov, (301) 286-7693

John R. Wilson, Aerospace Corp., john.r.wilson@aero.org, (310) 336-0102

Application area and specific scenario.

The ADGP effort will focus on employing advanced data management technologies to support the production, storage and distribution of massive data sets in a heterogeneous, distributed computing environment. The ADGP will operate with three nodes at the following locations:

- NASA GSFC
- NASA Ames
- Aerospace Corp., El Segundo, CA

The ADGP data system at GSFC will participate in both the ADGP and the CEOS data grid prototype, and afford the opportunity to implement functionality to manage and transfer massive data sets across multiple, intersecting data grids.

Technology exploration focused scenario.

The prototype will be implemented in four phases in order to investigate progressively more complex issues associated with data intensive computing environments. These phases are:

- Training phase
- Base-level phase
- Mass storage system (MSS) / Prototype demonstration phase.

Training Phase (0 to 6 months)

Training (NASA GSFC)	- Introduce development personnel to SRB/MCAT/DataCutter, Globus Toolkit: functions, operations and associated APIs - Establish single SRB-server that manages two disk storage systems, i.e. directly connected disk array and Storage Area Network (SAN), and provides workstations/terminals that support software development environment - DEVELOP METADATA SCHEMA FOR MODIS LEVEL 0/LEVEL 1 DATA - RECEIVE VIA DIGITAL MEDIA, CATALOG, STORE MODIS LEVEL 0 DATA/METADATA - DEVELOP BASIC APPLICATIONS EMPLOYING SRB/MCAT APIS, E.G. DATA INGEST, MCAT UPDATE

Base-level Phase. (6 to 12 months)

Base-level (All Sites)	•	Establish internal interfaces between/among all ADGP sites, i.e. message and data communications interfaces supporting grid operations and data transfers
Base-level (NASA GSFC)	•	Establish network/data communications (external) interface between GSFC site and GSFC DAAC
	•	Receive from GSFC DAAC, ingest, catalog, store MODIS Level 0 data/metadata
	•	Modify MODAPS Level 1 production software for execution in SRB/MCAT environment
	•	TRANSFER MODIS LEVEL 1 DATA/METADATA TO NASA AMES VIA GRID-ENABLED APPLICATIONS SOFTWARE
	•	PERFORM FOLLOWING OPERATIONS SIMULTANEOUSLY: LEVEL 0 DATA INGEST/STORAGE, LEVEL 1 DATA/METADATA PRODUCTION AND STORAGE, LEVEL 1 DATA TRANSFER (TO AMES) AND MCAT UPDATE
	•	MEASURE PERFORMANCE OF LEVEL 0 DATA INGEST, LEVEL 1 DATA PRODUCTION, SRB/MCAT OPERATIONS, LEVEL 1 DATA/METADATA TRANSFER
Base-level (NASA Ames)	•	Install and test SRB/MCAT/DataCutter in data system supporting ADGP
	•	Install metadata schema for MODIS Level 1 data
	•	Receive MODIS Level 1 data/metadata from GSFC via grid operations
	•	Ingest, catalog, store Level 1 data/metadata
	•	Measure performance of SRB/MCAT and data store operations
Base-level (Aerospace)	•	Develop, integrate and test applications software for 'science user' workstation, e.g. data visualization tools, related grid-enabled query and data retrieval applications
	•	Employing grid-enabled applications, retrieve from GSFC and ingest MODIS Level 1 data as input to science applications
	•	Employing grid-enabled applications, retrieve from NASA Ames and ingest MODIS Level 1 data as input to science applications
	•	Measure performance of grid-enabled applications and grid operations

Mass Storage/ADGP Demonstration Phases. (13 to 24 months)

Mass Storage System (NASA GSFC)	•	Modify base-level hardware configuration to implement high-availability data ingest, tape storage of MODIS Level 0 data. Reconfigure SRB/MCAT accordingly.
	•	Generate/transfer 50-100 terabytes of MODIS Level 1 data and associated metadata to NASA Ames
	•	Maintain MCAT
	•	Measure performance of Level 0 data/metadata ingest, Level 1 data production, SRB/MCAT/DataCutter operations, Level 1 data/metadata transfer
Mass Storage System (NASA	•	Install and test tertiary (tape) storage subsystem and reconfigure SRB/MCAT
	•	Receive, ingest, catalog, store 50-100 terabytes of MODIS Level 1 data/metadata
	•	Modify science application(s), e.g. climate model(s), for SRB/MCAT environment
	•	Retrieve, transfer MODIS Level 1 data/metadata in response to local/remote queries and I/O commands issued by science application(s)

Ames)	• Measure performance of Level 1 data/metadata ingest, local and remote access to Level 1 data/metadata, SRB/MCAT/DataCutter operations, data transfer, application execution
Mass Storage System (Aerospace)	• Initiate execution of science applications locally and at Ames, employing grid-enabled applications • Measure performance of grid-enabled applications and grid operations
ADGP Demonstration (All Sites)	• Develop and execute grid-enabled applications described in Section 3.5 • Measure performance of grid-enabled applications and grid operations

Grid locations.
The ADGP will operate with three nodes at the following locations:
- NASA GSFC
- NASA Ames
- Aerospace Corp., El Segundo, CA

The ADGP data system at GSFC will participate in both the ADGP and the CEOS data grid prototype, and afford the opportunity to implement functionality to manage and transfer massive data sets across multiple, intersecting data grids.

Confirmed Datasets.

None

Potential Datasets
- MODIS (Level 0)
- MODIS (Level 1)

Grid services.
ADGP applications will employ/evoke the following grid services:
- Resource registration
- Resource status monitoring
- Resource discovery, scheduling and reservation
- Grid application initiation
- Data transfer between nodes in the data grid
- Grid security, i.e. user certification and validation
- Grid event monitoring (**TBR**).

Applications and data. (TBR)
During the final demonstration of ADGP capabilities. A 'scientist' working from a workstation located in a facility of the Aerospace Corp., El Segundo, CA, will:
- Prepare and launch a grid-enabled application that:
  - Requests, schedules and reserves computing and storage resources at NASA Ames
  - Initiates the transfer of a subset of 'current' MODIS Level 1 data from GFSC to NASA Ames
  - Initiates execution of application software, e.g. a climate model, at NASA Ames. Application retrieves subsetted Level 1 data from local storage, employs it as input and returns final output to Aerospace Corp.
  - Replicates/transfers output to GFSC for storage

- Prepare and launch a grid-enabled application that:
    - Requests, schedules and reserves computing and storage resources at NASA Ames
    - Initiates execution of application software, e.g. a climate model, at NASA Ames that retrieves a subset of archived Level 1 data from the tertiary (tape) storage system at Ames, employs it as input to the application, and returns final output to Aerospace Corp.
    - Replicates/transfers output to GFSC for storage.